FORECASTING

Herman J. Bierens

Pennsylvania State University

September 11, 2002

1. Recursive best linear forecasting

Let Y_t be a covariance stationary time series process, with $E[Y_t] = 0$. The best linear h-step ahead forecast of Y_{t+h}, $h = 1,2,3,...,$ given the observations on $Y_t, Y_{t-1}, Y_{t-2}, ...,$ [1] is a linear function of Y_{t-j}, $j = 0,1,...,$ say:

$$\hat{Y}_{t+h|t} = \sum_{j=0}^{\infty} \gamma_{h,j} Y_{t-j}, \tag{1}$$

such that the mean-square forecast error

$$E\left(Y_{t+h} - \hat{Y}_{t+h|t}\right)^2 = E\left(Y_{t+h} - \sum_{j=0}^{\infty} \gamma_{h,j} Y_{t-j}\right)^2 \tag{2}$$

is minimal. Therefore, the coefficients $\gamma_{h,j}$ are such that the first-order conditions

$$E\left(Y_{t+h} - \sum_{j=0}^{\infty} \gamma_{h,j} Y_{t-j}\right) Y_{t-k} = 0 \ for \ k = 0,1,2,...,\tag{3}$$

are satisfied.

Note that we can write **(2)** and **(3)** in terms of the covariance function

$$f(m) = cov\left(Y_t, Y_{t-m}\right) = E[Y_t Y_{t-m}] \tag{4}$$

---

[1] In practice we do not observe the whole past of a time series, of course. What follows in this section is only a theoretical exercise. The practical aspects will be addressed in section 3.

(the last equality follows from the assumption that $E[Y_t] = 0$):

$$E\left(Y_{t+h} - \hat{Y}_{t+h|t}\right)^2 = f(0) - 2\sum_{j=0}^{\infty} \gamma_{h,j} f(h+j) + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \gamma_{h,i}\gamma_{h,j} f(|i-j|) \quad \textbf{(5)}$$

with first-order conditions:

$$f(h+k) = \sum_{j=0}^{\infty} \gamma_{h,j} f(|k-j|), \; k = 0, 1, 2, ... \quad \textbf{(6)}$$

Given the covariance function $f()$, we can in general solve the coefficients $\gamma_{h,j}$ uniquely from **(6)**.

Now consider the best linear one-step ahead forecast of Y_{t+2}:

$$\hat{Y}_{t+2|t+1} = \sum_{j=0}^{\infty} \gamma_{1,j} Y_{t+1-j} = \gamma_{1,0} Y_{t+1} + \sum_{j=0}^{\infty} \gamma_{1,j+1} Y_{t-j}. \quad \textbf{(7)}$$

This expression can be rewritten as

$$\hat{Y}_{t+2|t+1} = \gamma_{1,0}\left(Y_{t+1} - \hat{Y}_{t+1|t}\right) + \gamma_{1,0}\hat{Y}_{t+1|t} + \sum_{j=0}^{\infty} \gamma_{1,j+1} Y_{t-j}. \quad \textbf{(8)}$$

It follows from the first-order conditions that for $k = 0,1,2,....$

$$\begin{aligned} 0 &= E\left(Y_{t+2} - \hat{Y}_{t+2|t+1}\right)Y_{t-k} = E\left[-\gamma_{1,0}\left(Y_{t+1} - \hat{Y}_{t+1|t}\right) + Y_{t+2} - \gamma_{1,0}\hat{Y}_{t+1|t} - \sum_{j=0}^{\infty} \gamma_{1,j+1} Y_{t-j}\right]Y_{t-k} \\ &= -\gamma_{1,0} E\left(Y_{t+1} - \hat{Y}_{t+1|t}\right)Y_{t-k} + E\left[Y_{t+2} - \gamma_{1,0}\hat{Y}_{t+1|t} - \sum_{j=0}^{\infty} \gamma_{1,j+1} Y_{t-j}\right]Y_{t-k} \quad \textbf{(9)} \\ &= E\left[Y_{t+2} - \gamma_{1,0}\hat{Y}_{t+1|t} - \sum_{j=0}^{\infty} \gamma_{1,j+1} Y_{t-j}\right]Y_{t-k}, \end{aligned}$$

hence:

$$\hat{Y}_{t+2|t} = \gamma_{1,0}\hat{Y}_{t+1|t} + \sum_{j=0}^{\infty}\gamma_{1,j+1}Y_{t-j}. \quad \textbf{(10)}$$

More generally we have:

THEOREM 1. *Let* Y_t *be a covariance stationary time series process, with* $E[Y_t] = 0$. *Let* Y_{t-j} *be observable for j* = 0,1,2,.... *Replacing in the expression for the best linear one-step ahead forecast* $\hat{Y}_{t+h|t+h-1}$ *of* Y_{t+h}, *i.e.,*

$$\hat{Y}_{t+h|t+h-1} = \sum_{j=0}^{\infty}\gamma_{1,j}Y_{t+h-1-j} = \sum_{j=0}^{h-2}\gamma_{1,j}Y_{t+h-1-j} + \sum_{j=h-1}^{\infty}\gamma_{1,j}Y_{t+h-1-j} \quad \textbf{(11)}$$

the unobserved $Y_{t+h-1-j}$, $j = 0,....,h-2$, *by best linear forecasts* $\hat{Y}_{t+h-1-j|t}$, *respectively, yields the best linear h-step ahead forecast of* Y_{t+h}:

$$\hat{Y}_{t+h|t} = \sum_{j=0}^{h-2}\gamma_{1,j}\hat{Y}_{t+h-1-j|t} + \sum_{j=h-1}^{\infty}\gamma_{1,j}Y_{t+h-1-j}. \quad \textbf{(12)}$$

If $E[Y_t] = \mu \neq 0$, the best linear *h*-step ahead forecast takes the form

$$\hat{Y}_{t+h|t} = \delta_h + \sum_{j=0}^{\infty}\gamma_{h,j}Y_{t-j}. \quad \textbf{(13)}$$

Exercise 1: Show that

$$\delta_h = \left(1 - \sum_{j=0}^{\infty}\gamma_{h,j}\right)\mu \quad \textbf{(14)}$$

with the coefficients $\gamma_{h,j}$ determined by **(6)**, so that

$$\hat{Y}_{t+h|t} = \mu + \sum_{j=0}^{\infty} \gamma_{h,j}\left(Y_{t-j} - \mu\right). \tag{15}$$

The practical implication of this result is that in forecasting Y_{t+h} we may first forecast $Y_{t+h} - \mu$, using the result of Theorem 1, and then add μ to the forecast involved.

Exercise 2: Prove that:

THEOREM 2: *For the case* $E[Y_t] = \mu \neq 0$ *the result of Theorem 1 becomes*

$$\hat{Y}_{t+h|t} = \sum_{j=0}^{h-2} \gamma_{1,j}\hat{Y}_{t+h-1-j|t} + \sum_{j=h-1}^{\infty} \gamma_{1,j}Y_{t+h-1-j} + \left(1 - \sum_{j=0}^{\infty} \gamma_{1,j}\right)\mu. \tag{16}$$

2. *Forecasting with an* ARMA(p,q) *model*: *Theory*

Consider the ARMA(p,q) process

$$\begin{gathered} Y_t = \mu + u_t,\ \alpha(L)u_t = \beta(L)e_t, \\ \textit{where} \\ \alpha(L) = 1 - \sum_{j=1}^{p} \alpha_j L^j, \quad \beta(L) = 1 - \sum_{j=1}^{q} \beta_j L^j, \\ \alpha(z) = 0 \Rightarrow |z| > 1, \quad \beta(z) = 0 \Rightarrow |z| > 1, \\ e_t \textit{ is white noise}: E(e_t) = 0, \quad E(e_t^2) = \sigma^2 < \infty, \quad E(e_t e_{t-j}) = 0 \textit{ for } j \neq 0. \end{gathered} \tag{17}$$

Moreover, we have to assume that the lag polynomials $\alpha(L)$ and $\beta(L)$ do not have common roots (*Exercise 3*: Why?). Since the lag polynomial $\beta(L)$ is invertible, because all its roots are outside the unit circle, we can write this process as an AR(∞) process:

$$\gamma(L)(Y_t - \mu) = e_t, \textit{ where } \gamma(L) = \beta(L)^{-1}\alpha(L) = 1 - \sum_{j=0}^{\infty} \gamma_j L^{j+1}, \tag{18}$$

say. Note that $\gamma(z) = 0 \Rightarrow |z| > 1$. Thus:

$$Y_{t+1} = \delta + \sum_{j=0}^{\infty} \gamma_j Y_{t-j} + e_{t+1}, \quad where\ \delta = \left(1 - \sum_{j=0}^{\infty} \gamma_j\right)\mu. \qquad \textbf{(19)}$$

Consequently, the best linear one-step ahead forecast of Y_{t+1} is:

$$\hat{Y}_{t+1|t} = \delta + \sum_{j=0}^{\infty} \gamma_j Y_{t-j} = \mu + \sum_{j=0}^{\infty} \gamma_j \left(Y_{t-j} - \mu\right). \qquad \textbf{(20)}$$

(*Exercise 4*: Why?) Using Theorem 2, we can recursively find the best linear h-step ahead forecast of Y_{t+h} by

$$\hat{Y}_{t+h|t} = \sum_{j=0}^{h-2} \gamma_j \hat{Y}_{t+h-1-j|t} + \sum_{j=h-1}^{\infty} \gamma_j Y_{t+h-1-j} + \left(1 - \sum_{j=0}^{\infty} \gamma_j\right)\mu \qquad \textbf{(21)}$$

3. *Forecasting with an* ARMA(*p,q*) *model*: *Practice*

The practical problem with the above approach is three-fold: First, we usually do not observe the whole process Y_t, but only a finite number of Y_t's, say for $t = 1,...,n$. Second, p and q are unknown. We will address that problem later. Third, we do not observe the coefficients α_i, β_j directly. These coefficients have to be estimated. The latter can be done by maximum likelihood, but that requires further assumptions on the distribution of the white noise errors e_t.

An alternative approach is nonlinear least squares estimation, together with the assumption that $e_t = 0$ for $t < 1$, hence $u_t = 0$ for $t < 1$ and $Y_t = \mu$ for $t < 1$. The assumption $e_t = 0$ for $t < 1$ is asymptotically innocent: it does not affect the consistency or asymptotic normality of the parameter estimates. The least squares problem involved is:

$$\min_{\theta} \sum_{t=1}^{n} e_t(\theta)^2,$$
$$subject \;\; to$$
$$e_t(\theta) \; = \; \sum_{j=1}^{q} \beta_j I(t-j>0) e_{t-j}(\theta) \;+\; Y_t - \mu \;-\; \sum_{j=1}^{p} \alpha_j I(t-j>0)\left(Y_{t-j} - \mu\right), \;\; t \;=\; 1,..,n, \tag{22}$$
$$where \;\; \theta \; = \; (\mu, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q)^T,$$

with $I()$ the indicator function: I(true) = 1, I(false) = 0. Under some regularity conditions, in particular the condition that p and q are correctly specified, and the condition that the errors e_t are martingale differences: $E[e_t | e_{t-1}, e_{t-2}, e_{t-3}, ...] \; = \; 0$, it can be shown that the nonlinear least squares estimator $\hat{\theta} \; = \; (\hat{\mu}, \hat{\alpha}_1, ..., \hat{\alpha}_p, \hat{\beta}_1, ..., \hat{\beta}_q)^T$ is consistent and asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta} \; - \; \theta) \; \to \; N_{p+q+1}[0, \Omega_1^{-1}\Omega_2\Omega_1^{-1}] \;\; in \;\; distribution,$$
$$where$$
$$\Omega_1 \; = \; \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E\left[\left(\frac{\partial e_t(\theta)}{\partial \theta^T}\right)\left(\frac{\partial e_t(\theta)}{\partial \theta}\right)\right], \tag{23}$$
$$\Omega_2 \; = \; \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E\left[e_t(\theta)^2 \left(\frac{\partial e_t(\theta)}{\partial \theta^T}\right)\left(\frac{\partial e_t(\theta)}{\partial \theta}\right)\right].$$

Moreover, these two matrices can be consistently estimated by

$$\hat{\Omega}_1 \; = \; \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial e_t(\theta)}{\partial \theta^T}\right)\left(\frac{\partial e_t(\theta)}{\partial \theta}\right)\Bigg|_{\theta=\hat{\theta}}, \quad \hat{\Omega}_2 \; = \; \frac{1}{n} \sum_{t=1}^{n} e_t(\theta)^2 \left(\frac{\partial e_t(\theta)}{\partial \theta^T}\right)\left(\frac{\partial e_t(\theta)}{\partial \theta}\right)\Bigg|_{\theta=\hat{\theta}}, \tag{24}$$

respectively.

Once we have estimated the parameters α_i, β_j, we can compute the γ_j 's recursively, as follows. Observe from **(18)** that

$$e_t \; = \; u_t \; - \; \sum_{j=0}^{\infty} \gamma_j u_{t-1-j}. \tag{25}$$

If we set in **(25)** $u_t = -1$ for $t = -1$, $u_t = 0$ for $t \neq -1$, then

$$\begin{gathered} e_j = 0 \textit{ for } j < -1 \\ e_{-1} = u_{-1} = -1 \\ e_0 = u_0 - \gamma_0 u_{-1} = \gamma_0 \\ e_1 = u_1 - \gamma_0 u_0 - \gamma_1 u_{-1} = \gamma_1 \\ e_2 = u_2 - \gamma_0 u_2 - \gamma_1 u_0 - \gamma_2 u_{-1} = \gamma_2 \\ \ldots\ldots\ldots\ldots \\ e_t = \gamma_t,\ t \geq 0. \end{gathered} \tag{26}$$

But it follows from **(17)** that also

$$e_t = \sum_{j=1}^{q} \beta_j e_{t-j} + u_t - \sum_{j=1}^{p} \alpha_j u_{t-j}. \tag{27}$$

Thus if we set in **(27)**, $u_t = -1$ for $t = -1$, $u_t = 0$ for $t \neq -1$, $e_j = 0$ for $j < -1$, then $e_{-1} = -1$ and $e_t = \gamma_t$ for $t = 0,1,2,.....$ Therefore, the γ_j's can be solved recursively, on the basis of the nonlinear least squares estimation results, by:

$$\begin{gathered} \hat{\gamma}_{-1-j} = 0 \textit{ for } j > 0, \\ \hat{\gamma}_{-1} = -1, \\ \hat{\gamma}_0 = \hat{\alpha}_1 - \hat{\beta}_1, \\ \hat{\gamma}_j = \sum_{i=1}^{q} \hat{\beta}_i \hat{\gamma}_{j-i} + \hat{\alpha}_{j+1} \textit{ for } j = 2,....,p-1, \\ \hat{\gamma}_j = \sum_{i=1}^{q} \hat{\beta}_i \hat{\gamma}_{j-i} \textit{ for } j \geq p. \end{gathered} \tag{28}$$

Replacing in **(20)** the Y_{t-j} for $j \geq t$ by $\hat{\mu}$ and the other parameters by their estimates, yields the feasible best linear one-step ahead forecast

$$\tilde{Y}_{t+1|t} = \hat{\mu} + \sum_{j=0}^{t-1} \hat{\gamma}_j \left(Y_{t-j} - \hat{\mu}\right),$$

and replacing in **(21)** $Y_{t+h-1-j}$ for $j \geq t+h-1$ by $\hat{\mu}$ and the other parameters by their estimates, yields the recursive formula for the feasible best linear h-step ahead forecast:

$$\tilde{Y}_{t+h|t} = \hat{\mu} + \sum_{j=0}^{h-2} \hat{\gamma}_j \tilde{Y}_{t+h-1-j|t} + \sum_{j=h-1}^{t+h-2} \hat{\gamma}_j \left(Y_{t+h-1-j} - \hat{\mu} \right). \tag{30}$$

As to the choice of p and q, there are a few model selection tools on the market such as the Akaike information criterion. However, if forecasting is the goal, then the out-of-sample forecasting performance is a the best criterion. Thus, select a sub-sample $Y_1, \ldots\ldots, Y_m$, $m < n$, and estimate the parameters of the ARMA(p,q) model using the sub-sample only. Then choose p and q such that the sum of squared out-of-sample forecast errors,

$$\Sigma_{h=1}^{n-m} (Y_{m+h} - \tilde{Y}_{m+h|m})^2$$

is minimal. Once you have determined p and q, re-estimate the parameters using the whole sample $Y_1, \ldots\ldots, Y_n$,, and forecast Y_{n+h} by $\tilde{Y}_{m+h|n}$.